Air Quality Data

The United States Environmental Protection Agency maintains air quality data information for approximately 12,000 sites. Presently only about 4,000 sites are operational (7). Prior to 1972, air quality measurements were net undertaken on a large scale, and were often subject to considerable measurement errors. The EPA data are published annually in Air Quality Data - Annual Statistics. The air quality data used in the statistical analysis presented below is from this 1977 annual publication.

Air quality data was matched to individual data from the NAS Twins Registry by three digit zip code. The most disaggregated measure of air quality was found to be based on three digit zip codes. Five digit zip codes were not a useful basis for air quality data collection because the number of correspondences between air quality monitoring sites and five digit zip codes was minimal.

The data actually collected by three digit zip code included: maximum 24 hour measurement for total suspended particulate and sulfur dioxide; and type of monitoring station.

Frequently it was necessary to choose between a number of monitoring sites as representing air quality measurements for a given three digit zip code. The criteria by which such decisions were made were: (1) discard all sites for which measurements were discontinued before the end of the year 1977, (2) discard all sites which were not identified by type of monitoring station, (3) choose that site which measures the largest number of pollutants, (4) if two or more sites measure the same number of pollutants, choose the site which has operated the longest, (5) if a choice cannot be made, choose the site with the largest number of measurements for total suspended particulate and (6) if a choice still cannot be made, choose randomly. Note that these criteria were to be applied in sequence from first to sixth.

The importance of the monitoring station type is with respect to matching the air quality data to individual twin registry data. It was pointed out in discussing Figure 4 that residence and work history information was obtained by Q2 with reference to urban, suburban, or rural locations. Similarly, air quality monitoring stations are identified as being located in “center city”, “suburban”, or “rural”. Therefore, air quality data collection was based both on three digit zip codes and on the urban, suburban, rural classification. For each three digit zip code, the goal was to find an urban, suburban, and rural measurement. Unfortunately, this was not always possible. Finally, the actual combination of health data with air quality data has been accomplished by matching the most recent individual residence urban-suburban-rural location by three digit zip code and with the appropriate urban-suburban-rural three digit zip code air quality data.

Unfortunately, the various data sets apply to different points in time in that symptoms increase, and air pollution concentrations are measured in 1967, 1973, and 1977, respectively. In addition, there are difficulties in relating long term air pollutant exposures to individuals at the last location they have resided at. More than one half of the twins have resided since 1945 in two or more locations, and it is unlikely that ambient concentrations in the different locations would be comparable. A second qualification is that cumulative estimates of cigarette or alcohol consumption have not been calculated. In consequence, current non-smokers may have symptoms but have no current cigarette consumption.

STATISTICAL RESULTS

In this section, a reasonably meaningful sub-set of the statistical results are presented along with a partial interpretation of their meaning. The data set after calling out observations with incomplete data or unusable responses to questions ended up being between 7,892 and 7,908 in number. This represents slightly more than 50 percent of the original NAS twins data set. Most of the deletions were due to the inability to obtain matching zip codes between the living location of the twins and an air monitoring station. The bias resulting from this omission is not known. However, it can be anticipated that most of these omissions are of twins residing in suburban or rural locations without monitoring stations, in which case there are fewer observations on those exposed to lower ambient air pollutant concentrations. The effect is to give less dispersion to exposures and thereby insert an indeterminate bias on the estimated coefficients and make their significance less than would be the case.

Given the lack of dispersion in age, socio-economic class, and race we should also anticipate a bias downward in estimated effects of air pollution exposures as contrasted to the U.S. total population. The relative uniformity of the NAS twins sample reduces problems of bias associated with comparing non-homogeneous groups and unknown group differences but increases the liklihood that nothing will be detected connecting air pollution to symptoms of disease when in fact there is a connection.

With these qualifications in mind, we now turn to the actual results. In Table 3 are recorded the means and standard deviations of the variables examined. In Table 4, a raw correlation matrix of results is presented for all of the variables. There is very little correlation between most of the variables with two notable exceptions. There is substantial correlation between the various measures of nutrients and minerals consumed. For example, the raw correlation coefficient between sugar and unsaturated fatty acids consumption is .75. While the correlation coefficient between calcium and vitamin A consumption is .84. Relatively high correlations were also observed among symptoms, which might be anticipated in that severe chest pain is a form of chest pain ($r = 0.32$) and cough and shortness of breath may occur simultaneously, ($r = 0.19$). For the remainder of variables, there is little or no raw correlation which would be expected of a relatively homogeneous data set of 8,000 observations.

Evaluation of Statistical Results from Regressions

After some preliminary experiments with the NAS twins data set, several conclusions emerged. First, there was no effect of the Twins on the estimated relationship between prevalence of a symptom and exposure to air pollution. Thus, there appears to be no discernible "genetic" effect at least in the sample analyzed. Second, a variable reflecting zygosity of twins was never even marginally significant. However, much more detailed statistical comparisons would need to be made in order to rule out the possible

TABLE 6.2 DEFINITION OF VARIABLES*

TWNO	Twin Number	Number of twin
CHPN	Chest Pain	Whether the individual experienced chest pain in 1967, (yes or no)
SHBR	Shortness of Breath	As measured by self-reported statement as to whether it was encountered when walking with friends, in 1967
COGH	Cough	Whether or not the individual regularly or for extended periods of time had a cough before or during 1967
SVCP	Severe Chest Pain	Whether the individual experienced severe chest pain lasting one half hour or more in 1967, (yes or no)
CORN	Coronary	Whether or not the individual had suffered a heart attack before or during 1967
ASTM	Asthma	Whether the individual had asthma before or during 1967, (yes or no)
RHMF	Rheumatic Fever	Whether or not the individual had rheumatic fever or rheumatic heart disease during or before 1967
DIET	Diet	Whether the individual undertook a diet for excess weight before or during 1967, (yes or no)
SMKN	Smoking	Cigarette consumption (packs per day) where conversions are used for cigars and pipe smokers before or during 1967
DRNK	Drunk	How often did the individual drink at least one pint of liquor or two bottles of wine or four quarts of beer at one occasion in 1967
INTX	Intoxication	Whether or not the individual becomes intoxicated daily, in 1967
CTRM	Cigarette Tar	Tar from cigarettes in milligrams per year, in 1967
LIQR	Liquor	Alcohol consumption, beer, wine, and spirits converted to ethanol equivalents in oz. per year, in 1967
HGHT	Height	Height in inches, in 1967
WGHT	Weight	Weight, in 1967

TABLE 6.2 (continued)

WT25	Weight at Age 25	Weight at age 25
BRTH	Birth	Year of birth
EARN	Earnings	Family earnings in 1973 (dollars)
TSPM	Max imum Total Sus- ' pended Parti- culates	Maximum 24 hour concentration in 1977, in $\mu g/m^3$
SO_2M	Maximum Sulfur Dioxide	Maximum 24 hour concentration in 1977, in $\mu g/m^3$
ZYGT	Zygosity	Classified as either monozygotic for identical twins and dizygotic for fraternal twins
STFT	Saturated Fatty Acids	Grams per year, in 1967
SUGR	Sugar	Grams per year, in 1967
FIBR	Fiber	Grams per year, in 1967
USFT	Unsatruated fatty acids	Grams per year, in 1967
NTRS	Nitrosamines	µg per year, in 1967
IRON	Iron	mg per year, in 1967
CALC	Calcium	mg per year, in 1967
THMN	Thiamin	mg per year, in 1967
NIAC	Niacin	mg per year, in 1967
VITA	Vitamin A	lu per year, in 1967
FATS	Fats	Grams per year, in 1967
PROT	Protein	Grams per year, in 1967
RIBF	Riboflavin	mg per year, in 1967

*Tables documenting conversions for food intake variables are reported in Appendix 1.

TABLE 6.3 MEANS AND STANDARD DEVIATIONS OF VARIABLES

Variable	Mean	Standard Deviation
CHPN	.24861	.43920
SHBR	.07145	.41695
COGH	.11292	.33212
SVCP	.04906	.21602
CORN	.11596	.85695
ASTM	.12355	.60376
RHMF	.03541	.18482
DIET	.22129	.41514
SMKN	.60255	.51997
DRNK	.85559	.38287
INTX	3.0567	14.208
CTRM	134.87	217.42
LIQR	425.37	643.76
HGHT	69.783	2.5466
WGHT	172.14	22.056
WT25	158.81	20.808
BRTH	22.956	2.9229
EARN	6.1792	11.687
TSPM	129.54	144.29
SO_2M	49.594	88.189
ZYGT	1.5622	.55714
STFT	7.5156	2.8017
SUGR	51.575	13.022
FIBR	.82409	.28507
US FT	8.6647	3.9050
NTRS	.07108	.06183
IRON	2.3021	.71629
CALC	.25948	.01400
THMN	.34392	.14039
NIAC	2.9235	.86349
VITA	.46757	.19714
FATS	17.436	6.7949
PROT	18.606	4.9021
RIBF	.57155	.08897

TABLE 6.4 CORRELATION MATRIX

	RIBF	PROT	FATS	VITA	NIAC	THMN	CALC	IRON	NTRS	USFT	FIBR	SUGR	STFT	ZYGT	SO_2M	TSPH	EARN	BRTW	WT25	WGHT	HGHT	LIQR	CTRM	INTX	DRNK	SMKN	DIET	RHHF	ASTH	CORN	SVCP	COGH	SHBR	CHPN	TWNO
TWNO	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.00	0.01	-0.00	0 00	0.01	0.00	0.01	-0.02	-0.03	-0.01	0.01	-0.00	-0.01	-0.04	-0.02	-0.01	-0.03	0.00	0.(31	-0.01	0.01	0.00	-0.01	-0.01	0.00	I. (MI
CHPN	-1.01	-0.01	-0.01	-0.01	-0.01	-0.02	0.00	-0.01	-0.02	-0.00	-0.02	0.01	-0.00	-0.00	0.02	-0.00	-0.03	0.00	-0. 05	-0.04	-0.02	0.05	0.06	0.04	0.02	0.04	0.04	0.01	0.07	0.20	0.)2	0.22	0.19	1.00	
SHBR	0.01	0.01	0.00	0.01	0.01	0.01	0.00	0.01	-0.02	0.01	-0.02	-0.01	0.01	0.01	-0.02	0.01	-0.04	-0.01	-0.01	0.01	-0.01	0.0S	0.08	0.07	-0.02	0.05	0.02	0.00	0.06	0.29	0.18	0.19	1.00		
COGH	-0.01	-0.01	-0.01	-0.01	0.01	-0.00	-0.01	-0.01	-0.01	-0.00	'0.06	0.01	-0.01	0.02	-0. ON	0.02	-0.04	-0.01	-0.03	-0.04	-0.03	0.15	0.27	0.01	0.07	0.17	-0.03	-0.01	-0.06	0.04	0.12	1.00			
SVCP	-0.01	-0.00	-0.01	-0.01	-0.01	-0.02	-0.01	-0.01	-0.02	-0.00	-0.01	-0.01	-0.00	0.01	0.02	0.01	-0.02	-0.01	-0.03	-0.03	-0.02	0.00	0.05	0,02	-0. W	0.02	0.05	-0.01	0.00	0.15	1.00				
CORN	-0.05	-0.04	-0.04	-0.04	-0.04	-0.05	-0.04	-0.05	-0.02	-0.04	-0.02	-0.01	-0.03	0.01	-0.01	0.03	-0.03	-0.05	0.01	-0.02	-0.02	0.01	0.03	0.02	-0.02	0.00	0.08	0.01	0.01	1.00					
ASTH	0.01	0.02	0.02	0.01	0.02	0.02	0.01	0.02	0.03	0.01	43.00	-0.00	0.01	-0.00	0.01	0.00	0.02	-0.01	-0.01	-0.00	0.03	0.01	0.00	0.02	-0.00	-0.01	0.01	0.00	1.00						
RHHF	-0.02	-0.02	-n. .01	-0.02	-0.01	-0.03	-0.01	-0.03	-0.02	-0.01	-0.01	0.00	-0.01	0.01	-0.01	-0.01	0.02	0.02	0.00	0.00	0.01	0.02	-0.01	-0.012	-0.00	0.01	0.02	1.00							
DIET	-0.05	-0.04	-0.05	-0.03	-0.09	-0.07	-0.07	-0.06	-0.00	-0.05	-0.06	-0.10	-0.04	0.00	0.03	0.04	0.09	'0.02	0.21	0.29	-0.00	0.02	-0.08	0.00	0.01	-0.07	1.00								
SMKN	0.01	0.02	-0.00	-0.01	0.05	0.03	-0.00	0.02	-0.00	-0.00	0.01	-0.00	0.00	0.00	-0.00	0.00	-0.04	0.03	-0.01	-0.07	-0.00	0.17	0.49	0.08	0.21	1.00									
DRNK	-0.10	-0.08	-0.10	-0.09	-0.06	-0.09	-0.09	-0.09	-0.04	-0.10	-0.01	-0.07	-0.09	-0.02	0.03	0.01	0.05	0.02	0.00	-0.00	0.02	0.26	0.13	0.09	1.00										
INTX	-0 02	0.01	-0.01	-n. 02	-0.00	-0.01	-0.02	-0.01	-0.01	-0.02	-0.06	-0.04	-0.01	0.01	-0.00	-0.02	-0.01	0.03	-0.01	-0. 00	-0. 01	0.32	0.14	1.00											
CTRM	-0.01	0.00	-0.01	-0.02	0.04	0.02	-0.03	0.00	0.00	-0.02	-0.04	-0.03	-0.01	0.01	-0.00	0.01	-0.07	0.05	-0.02	-0.07	-0.00	0.21	1.00												
LIQR	-0.11	-0. 07	-0.11	-0.10	-0.03	-0.05	-0.13	-0.07	-0.02	-0.12	'0.05	00.14	-0.10	0.00	0.02	-0.02	0.02	0.01	0.01	0.05	0.06	1.00													
HGHT	0.02	0.03	0.02	0.01	0.03	0.04	-0.01	0.03	0.01	1.01	0.01	-0.01	0.01	0.06	-0.(51	0.01	0.05	0.05	0.51	0.53	1.00														
WGHT	0.01	0.01	0.(I1	N.(11	-0.01	0.01	-0.00	0.01	0.03	0.01	-13.02	-0.01	0.01	0.03	-0.01	0.01	0.04	0.01	0.73	1.00															
WT25	-0.(32	-0.01	-0. 00	0.00	-0.03	-0.01	-0.02	-0.01	0.02	-0.02	'0.0²	-0.05	-0.01	0.05	-0.02	0.01	0.04	0.05	1.00																
BRTW	-0.01	-0.01	-0.01	-0.02	0.00	0.01	-0.02	-0.01	0.07	-0.01	0.02	0.01	-0.01	-0.03	0.01	0.01	-0.00	1.00																	
EARN	-0.01	0.00	-0.00	-0.01	-0.01	-0.01	-0.02	-0.00	-0.00	-0.01	0.01	-0.03	0.00	-0.06	-0.00	0.02	1.00																		
TSPH	-0.01	-0.00	-0.00	-0.02	0.00	-0.00	-0.02	-0.01	0.00	-0.01	-0.00	-0.02	-0.00	-0.02	0.22	1.00																			
SO_2M	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.01	-0.02	-0.01	-0.02	0.01	-0.00	-0.02	-0.00	1.00																				
ZYGT	0.01	0.01	0.02	0.01	0.01	0.01	0.02	0.01	-0.01	0.02	-0.01	0.01	0.01	1.00																					
STFT	0.90	0.90	0.93	0.64	0.76	0.59	0.80	0.86	0.21	0.9?	0.20	0.>1	1.00																						
SUGR	0.68	0.50	0.62	0.40	0.43	0.43	0.34	0.49	0.1S	0.75	0.40	1.00																							
FIBR	0.28	0.29	0.21	0.19	0.39	0.28	0.31	0.36	0.12	0.19	1.00																								
USFT	0.87	0.86	0.95	0.66	0.71	0.58	0.87	0.83	0.20	1.00																									
NTRS	0.33	0.39	0.42	0.14	0.61	0.?6	0.21	0.35	1.00																										
IRON	0.92	0.98	0.89	0.79	0.81	0.79	0.80	1.00																											
CALC	0..97	0.79	0.84	0.84	0.56	0.51	1.00																												
THMN	0.)0	0.79	0.74	0.46	0.81	1.00																													
NIAC	0.70	0.239	0.81	0.40	1.00																														
VITA	0.85	0.72	0.67	1.00																															
FATS	0.86	0.93	1.00																																
PROT	0.91	1.00																																	
RIBF	1.00																																		

presence of a genetic effect. Since no "genetic" effect was observed, the researchers decided to "pool" the usable twin observations for further statistical analysis. Third, ordinary least squares and probit statistical computations were made on the same data and no difference was observed in estimated coefficients or their standard errors. In consequence, statistical estimated concentrated almost exclusively on application of the ordinary least squares technique. Finally, it was observed that using a randomly drawn sample of twins to estimate the coefficients (of about 5 percent of the population) yielded coefficients in another. This suggests that for prediction purposes and accuracy, the entire population should be used for estimation purposes.

With approximately 8,000 unique and usable observations, it can be expected that R^2's will be relatively low and that was what was observed uniformly throughout the results.

In Table 5, are recorded the four variants of the regression equation for chest pain. The second equation is the same as the first except intake of sugar is added. For the third variant saturated fats is added, and in the fourth variant, vitamins, proteins, and minerals are added. Across the four variants, none of the independent variables' coefficients or "t" statistics changes very much. And the R^2's are uniformly low. The statistically significant variables are smoking, liquor consumption, but not heavy drinking, earnings, sugar intake, and to a lesser extent, maximum 24 hour concentrations of $S0_2$. As would be expected, smoking contributes to increased levels of chest pain (8). The most common mechanism would be smoke ingestion requiring more inspiration/expiration for the same level of oxygen and thereby greater requirements on the heart for pumping. Greater daily consumption of alcohol stresses the cardiovascular system so it is expected that this would have a positive effect on the incidence of chest pain (9). Birthdate or age has no impact, but this is to be expected given the sample age only ranges from 41 to 51 years. Earnings have a significant negative effect on chest pain. In this equation, earnings probably reflect education and knowledge of diseases and the demand for medical services plus other socio-economic effects. Thus, no economic interpretation (in demand and supply terms) can be made of the earnings coefficient. Finally, while the TSP coefficient is insignificant, the $S0_2$ coefficient is significant at the 95 percent confidence level, and remains stable in magnitude across the four variants of the regression. The coefficient indicates a one ten thousandth increase in the probability of chest pain given a 1 $\mu g/m^3$ increase in maximum average 24 hour concentrations of $S0_2$.

Table 6 contains the estimates for four variants depending on dietary specifications for the symptom, severe chest pain. Again, as with chest pain, smoking and whether the individual had dieted were statistically significant at the 97.5 percent level. Neither air pollution variables were significant across the four variants. Earnings again were negatively significant at the 95.5 percent level. It is curious that $S0_2$ would be significant for chest pain but not for severe chest pain. However, the severe chest pain variable is described as one that lasts one half hour or longer which may not adequately reflect the potential chronic effects of either $S0_2$ or TSp.

TABLE 6.5 ALTERNATIVE ORDINARY LEAST SQUARES RIXRESS1ONS WITH CNEST PAIN AS THE DEPENDENT VARIABLE. t Statistics ARE IN PARENTHESES

Dependent Variable and Regression #	Independent Variables DIET	SMKN	DRNK	LIQR	BRTH	EARN	TSPM	SO_2M	STFT	SUGR	FIBR	USFT
CHPN I	.0512 (4.285)	.0273 (2.787)	.0080 (.589)	.00003 (3.344)	-.00002 (-.010)	-.0013 (-2.947)	-.00002 (-.619)	.0001 (1.772)	-			
CHPN II	.0535 (4.452)	.0269 (2.743)	.0091 (.669)	.00003 (3.568)	-.0001 (-.036)	-.0012 (2.907)	-.00002 (-.580)	.0001 (1.756)	-	.0007 (1.948)	-	
CHPN III	.0540 (4.494)	.0277 (2.712)	.0088 (0.644)	.00003 (3.549)	.00001 (.004)	-.0012 (-2.775)	-.00002 (-.504)	-.0001 (1.719)	-.0071 (-. 768)	-.0018 (2.135)	-.0503 (-2.172)	.0059 (.644)
CHPN IV	.0522 (4.285)	.0280 (2.845)	.0070 (.516)	.00003 (3.576)	.0001 (.030)	-.0012 (-2.885)	-.00002 (-.539)	.0001 (1.699)	-			

	NTRS	IRON	CALC	THMN	NIAC	VITA	FATS	PROT	RIBF	CONSTANT	R^2	SSR DF
CHPN I										.2085 (5.108)	.0068	1515 7899
CHPN II										.169 (3.692)	.0073	1514 7898
CHPN III							-.0052 -1.25	.0043 (1.052)	-	.1659 (3. 000)	.0085	1512 7893
CHPN IV	-.1215 (-. 762)	.0188 (.209)	2.0581 (2.354)	-.0057 (-.050)	-.0237 (-.484)	-.1865 (-1.068)	-	-	.1232 (.646)	-.275 (-1.558)	.0084	1512 7892

Chest pain and severe chest pain symptoms are uniformly higher in individuals who have reported the necessity of dieting. This finding is collaborated by extensive medical research on the effect of excess weight on the liklihood of heart attacks and other cardiovascular problems (10).

In Table 7 are recorded a sample of the regression results obtained for the occurence of coronary heart attacks. The variable reflecting the need to diet is again highly positively significant. Smoking is less significant but still positive. Consumption of alcohol has a marginally significant effect while excessive drinking seems to have a negatively significant effect. Family earnings has the anticipated negative effect on the occurence of coronary heart attack. Of the air pollutant variables, TSP has a positive and highly significant impact on coronary heart attack. Alternatively, SO_2 is negatively related to coronary heart attack but the coefficient is only marginally significant. The consumption of more starches, fats, and nitrosamines has an apparent positive effect on heart attacks and protein a negative effect. Conceptually, from these regressions one could compare the effects of consumption of certain foods with suspended particulate as to relative effects on the prevalence of coronary heart attacks. That will not be done here because of the experimental nature of these results and the need for additional replication before the results can be accepted.

In Tables 8 and 9 are a sample of regression results for two respiratory symptoms, the presence of cough and shortness of breath. In both cases, TSP had a significant impact on their occurence, while SO_2 had a negative impact. For the presence of cough, smoking, liquor consumption, sugar intake, and TSP had highly significant positive effects. The need for dieting, family earnings, and fiber consumption had a negative impact. For shortness of breath, the need to diet, smoking, liquor consumption, and TSP had positive and significant effects on its incidence.

In Table 10, the “t” statistics are contrasted for the various symptoms and air pollutant variables. As was noted before, these do not vary greatly when dietary variables are included. Maximum average concentrations of TSP have a strong connection to the presence of three symptoms, coronary heart attack, cough, and with less significance, shortness of breath. Maximum average 24 hour concentration of SO_2 has a positive connection with the occurence of chest pain but a significant negative connection with coronary heart attack and shortness of breath. This anomolous result cannot be readily explained. However, SO_2 concentrations are higher in heavy manufacturing-industrial areas where workers doing physical labor may be in relatively better physical condition due to exercise. In consequence, the answer to the shortness of breath question might be biased since it references walking on level ground with other people. Healthier individuals resulting from physical exercise at work may not respond to the shortness of breath question even though there may be some respiratory impairment.

Table 11 presents elasticities of the incidence rate of a symptom with respect to air pollution. These elasticities represent point estimates of elasticity about the mean. They derived via the following formula

TABLE 6.6 ALTERNATIVE ORDINARY LEAST SQUARES REGRESSIONS WITH SEVERE CHEST PAIN AS THE DEPENDENT VARIABLE. t STATISTICS ARE IN PARENTHESES

Dependent Variable and Regression #	Independent Variables											
	DIET	SMKN	DRNK	LIQR	BRTH	EARN	TSPM	SO_2M	STFT	SUGR	FIBR	USFT
SVCP I	.0290 (4.928)	.0120 (2.496)	-.0059 (-.878)	.000001 (.195)	-.0008 (-1.016)	-.0003 (-1.652)	.00001 (.360)	-.000005 (-.168)	-	-	-	
SVCP 11	.0290 (4.895)	.0121 (2.498)	-.0059 (-.882)	.000001 (.180)	-.0008 (-1.015)	-.0003 (-1.654)	.00001 (.358)	-.000005 (-.167)	-	-.00002 (-.1048)	-	
SVCP III	.0289 (4.865)	.0119 (2.457)	-.0059 (-.876)	.000001 (.167)	-.0008 (-.920)	-.0003 (-1.621)	.00001 (.384)	-.00001 (-.178)	-.0017 (-.366)	-.0002 (-.362)	-.0063 (-.555)	.0054 (1.210)
SVCP IV	.0296 (4-931)	.0120 (2.481)	-.0067 (-.995)	.000001 (.151)	-.0008 (-.961)	-.0003 (-1.629)	.00001 (.381)	-.00006 (-.200)	-			
	NTRS	IRON	CALC	THMN	N IAC	VITA	FATS	PROT	RIBF	CONSTANT	R^2	SSR DF
SVCP I										.0610 (3.035)	.0041	367 7899
SVCP II										.0621 (2.760)	.0041	367 7898
SVCP III							-.0034 (-1.680)	.0017 (.825)	-	.0662 (2.431)	.0046	367 7893
SVCP IV	-.0528 (-.673)	-.0104 (-.234)	.1766 (.410)	-.0315 (-.560)	.0107 (.445)	.0071 (.082)	-		-.0046 (-.049)	.0213 (.245)	.0048	368 7892

TABLE 6.7 ALTERNATIVE ORDINARY LEAST SQUARES WITH THE INCIDENCE OF CORONARY HEART ATTACK AS THE DEPENDENT VARIABLE. c STATISTICS ARE IN PARENTHESES

Dependent Variable and Regression #	Independent Variables DIET	SMRN	DRNK	LIQR	BRTH	EARN	TSPM	SO_2M	STFT	SUGR	FIBR	USFT
CORN I	.1686 (7.248)	.0183 (.962)	-.0.427 (-1.617)	.00002 (1.287)	-.0156 (-4.758)	-.2740 (-3.316)	.0002 (2.763)	-.0001 (-1.267)	-			
CORN 11	. 164(I (7.017)	.0192 (1.009)	-.0450 (-1.701)	-00002 (1.006)	-.0155 (-4.722)	-.0028 (-3.3S9)	.0002 (2.721)	-.0001 (-1.250)	-	-.0016 (-2.095>	-	
CORN III	.1607 (6.847)	.0231 (1.211)	-.0513 (-1.939)	.00002 (1.066)	-.0158 (-4.802)	-.0028 (-3.354)	.0002 (2.688)	-.0001 (-1.297)	.0270 (1.500)	-.0019 (-1.165)	.0469 (1.042)	-.0433 (-.807)
CORN IV	.1551 (6. 550)	.0250 (1.309)	-.0555 (-2.091)	.00002 (1.031)	-.0158 (-4.826)	-.0028 (-3.423)	.0002 (2.729)	-.0001 (1,335)	-			

	NTRS	IRON	CALC	THMN	NIAC	VITA	FATS	PROT	RIBF	CONSTANT	R^2	SSR DF
CORN I										.4529 (5.701)	.0120	5737 7899
COKN II										.5364 (6.036)	.0126	5734 7898
CORN 111			, -				.0103 (1.281)	-.0227 (-2.845)	-	.6911 (6.425)	.0143	5724 7893
CORN Iv	.5283 (1.70s)	.2718 (1.548)	1.2481 (.734)	-.4602 (-2.073)	-.1206 (-1.268)	-.4820 (-1.420)	-		-.5041 (-1.360)	.5089 (1.483)	.0149	5120

TABLE 6.8 ALTERNATIVE ORDINARY LEAST SQUARES REGRESSIONS WITH COUGH AS THE DEPENDENT VARIABLE. t STATISTICS ARE IN PARENTHESES

Dependent Variable and Regression #	Independent Variables											
	DIET	SMKN	DRNK	LIQR	BRTH	EARN	TSPM	SO*M	STFT	SUGR	FIBR	USFT
coat I	-.0166 (-1.873)	.0952 (13.873)	.0097 (.966)	.99996 (10.366)	-.0019 (-1.537)	-.0010 (-3.306)	.00006 (2.229)	-.00004 (-.896)	-	-	-	-
COGH 11	-.0163 (-1.832)	.0951 (13.106)	.0098 (.979)	.00006 (10.322)	-.0019 (-1.541)	-.0010 (-3.298)	.00006 (2.235)	-.00004 (-.8991)	-	.00009 (.340)	-	-
COGH 111	-.0158 (-1.771)	.0950 (13.021)	.0121 (1.201)	.00006 (10.348)	-.0018 (-1.438)	-.0010 (-3.096)	.00006 (2.361)	-.00004 (-.865)	-.0119 (-1.739)	.0013 (2.108)	-.0944 (-5.513)	.0051 (.758)
COGH Iv	-.0102 (-1.132)	.0931 (12.792)	.0103 (1.024)	.00006 (10.308)	-.0018 (-1.474)	-.0009 (-3.012)	.00006 (2.350)	-.00004 (-.963)	-	-		

	NTRS	IRON	CALC	THMN	NIAC	VITA	FATS	PROT	RIBF	CONSTANT	R^2	SSR DF
COGH I										.0697 (2.305)	.0469	831 7899
COGH II										.0646 (1.908)	.0469	831 7898
COGH III							-.0024 (-.776)	.0053 (1.822)	-	.0571 (1.396)	.0512	828 7893
COGH Iv	-.1781 (-1.510)	-.2631 (-3.935)	-.3818 (-.590)	.0989 (1.17)	.1299 (3.585)	.4264 (3.298)	-		.2604 (1.844)	.0218 (.167)	.0490	829 7892

TABLE 6.9 ALTERNATIVE ORDINARY LEAST SQUARES REGRESSIONS WITH SHORTNESS OF BREATH AS THE DEPENDENT VARIABLE. t STATISTICS ARE IN PARENTHESES

Dependent Variable and Regression #	Independent Variables DIET	SMKN	DRNK	LIQR	BRTH	EARN	TSPM	SO_2M	ST FT	SUCR	FIBR	USFT
SHBR I	.0254 (2.240)	.0366 (3.937)	-.0421 (-3.276)	.00004 (4.8012)	-.0010 (-.613)	-.0013 (-3.168)	.00004 (1.162)	-.0001 (-2.195	-			
SHBR II	.0255 (2.234)	.0366 (3.935)	-.0421 (-3.270)	.00004 (4.768)	-.0010 (-.614)	-.0013 (-3.166)	.00004 (1.163)	-.0001 (-2.195)	-	.00002 (.054)	, -	
SHBR 111	.0256 (2.242)	.0350 (3.761)	-.0400 (-3.099)	.00004 (4.689)	-.0007 (-.410)	-.0013 (-3.129)	.00004 (1.208)	-.0001 (-2.169)	-.0088 (-.998)	-.0003 (-.377)"	-.0411 (-1.873)	.1769 (2.044)
SHBR IV	.0295 (2.554)	.0343 (3.681)	-.039'1 (-3.038)	.00004 (4.741)	-.0007 (-.468)	-.0013 (-3.111)	.00004 (1.139)	-.0001 (-2.149	-			

	NTRS	IRON	CALC	THMN	NIAC	VITA	FATS	PROT	RIBF	CONSTANT	R^2	SSR DF
SHBR I										.0957 (2.470)	.0082	1363 7899
SHBR II										.09 (2.183)	.0082	1363 7898
SHBR 111							-.0123 (-3.140)	.0106 (2.714)	-	.0669 (1.276)	.0103	1360 7893
SHBR IV	-.4710 (-3.117)	-.0670 (-.782)	-.0787 (-1.31)1)	.1877 (1..734)	.0283 (.609)	.0931 (.562)	-		.2437 (1.348)	.2258 (1.350)	.0102	1361 7892

TABLE 6.10 "t" STATISTICS ON AIR POLLUTION COEFFICIENTS*, SELECTED REGRESSIONS, NAS TWINS DATA SET

Symptom	Maximum Average 24 hour Concentrations** SO_2	TSP
Cardiovascular System		
Chest Pain	1.77[a]	-0.62
Severe Chest Pain	-0.17	0.36
Coronary Heart Attack	-1.27[d]	2.76[b]
Respiratory System		
Cough	-0.90	2.23[c]
Shortness of Breath	-2.19[c]	1.16[d]

*With nearly 8,000 observations, the "t" distribution approaches the normal distribution

[a]Significant at the 96% confidence level

[b]Significant at the 99.6% confidence level

[c]Significant at the 98% confidence level

[d]Significant at the 87% confidence level

**The simple correlation coefficient between TSP and SO_2 is .22.

TABLE 6.11 ELASTICITIES OF THE INCIDENCE RATE OF A SYMPTOM WITH RESPECT TO AIR POLLUTION*

Dependent Variable	Independent Variables Maximum 24 hour average concentration SO_2	 TSP
Chest Pain	1.995	-1.04
Severe Chest Pain	-0.505	2.64
Coronary Heart Attack	-5.988	21.23
cough	-1.757	6.88
Shortness of Breath	-8.329	7.25

*Elasticities are derived from coefficients in Equation 1 for all dependent variables at the mean values of the dependent and independent variables. This number represents the percent change in the probability of occurance of the symptom depicted by the dependent variable as a result of a 1 percent change in the independent variable

$$\text{Elasticity} = \frac{\text{change in the dependent variable}}{\text{change in the independent variable}} \cdot \frac{\text{mean of the independent variable}}{\text{mean of the dependent variable}}$$

Note however, that the first ratio on the right hand side of the above formula is simply the coefficient in the regression equation on the variable in question. This procedure allows the researcher to express results in percentage terms which are independent of the units used.

Care should be taken when interpreting the elasticities presented in Table 11. One should remember that the dependent variable is a probability. In this context, elasticities in the table represent the percentage change in the <u>probability</u> of the occurence of the event depicted by the dependent variable as a result of a one percent change in the independent variable. For example, if the maximum 24 hour average concentration of total suspended particulate increases by one percent then there will be a corresponding 21.23 percent change in probability of experiencing a coronary heart attack. However, the initial probability of a coronary heart attack (incidence) was slightly less than 12 percent in the sample. These values range for SO_2 from a low of -8.33 to a high of 2.00. Corresponding values for TSP range from a low of -1.04 to a high of 21.23.

What can be tentatively concluded from these experimental results? First, there appears to be a statistically significant connection between ambient concentration of total suspended particulate and several disease symptoms associated with both the respiratory and cardiovascular systems. Of particular importance is a strong and apparently replicative relationship between the incidence of coronary heart attacks and TSP. The evidence on concentrations of SO_2 and symptoms is much less clear. SO_2 is positively related to the self-reported occurence of chest pain. However, from these statistical results, SO_2 is negatively related to severe chest pain, coronary heart attack, cough, and shortness of breath. These findings should raise questions as to the reliability of self-reported data and the appropriateness of the questions themselves across diverse socio-economic groups.

Finally, regression equations were run omitting in sequence the SO_2 variable or the TSP variable. The omission of one of the air pollution variables had <u>no</u> influence on the magnitude, sign, or statistical significance of the included air pollutant variable. This lead us to the conclusion that the estimates reported in Tables 5 through 9 are relatively robust with regard to magnitude and sign.

ECONOMIC COSTS FROM POLLUTION

Lave and Seskin's (11) famous study, published in 1977, was one of the first to examine the statistical relationship between air pollution and health. They estimated the effects of air pollution, i.e., sulfur oxides and total suspended particulate, on the total mortality rate. Using the foregone earnings approach, they estimated benefits of pollution abatement via the reduction in the mortality rate. Lave and Seskin did not incorporate the relationship between air pollution and symptoms. Their approach focused on the direct relationship between air pollution and death.

Several other studies have been performed which relate air pollution and health. Most of these studies use mortality or morbidity rates as measures of health. For example, Crocker *et.al.* (12), 1979, use the mortality rate for pneumonia, influenza, emphysema, bronchitis and early infant disease as well as the total mortality rate for dependent variables. They used a variety of different air pollution measures as explanatory variables, concluding that only particulate and sulfur dioxide have statistically significant effects on health. Liu/Yu (13), 1979, utilized total mortality rates and the morbidity rate for bronchitis as health measures. They chose to use total suspended particulate and sulfur dioxides as pollution variables. Using both linear and non-linear models, they found that $S0_2$ and TSP have significant effects upon mortality and morbidity rates.

In contrast, this study focuses on the chain of events which link air pollution to the cost of increased symptoms due to air pollution. This methodology represents a substantial departure from that used in earlier studies.

Regression analyses, reported on earlier, were used to analyze the relationship between the occurence of a symptom and the factors affecting the symptom. Therefore, where Lave and Seskin use the mortality rate as the dependent variable, this report uses the occurence of a symptom such as cough, shortness of breath, etc. Coefficients on the independent variables give the change in the probability of a symptom given a unit change in a factor affecting the symptom.

Emphasis of this study is placed on the derivation of estimates of the reduction in costs of disease incurred when air pollution is reduced. The first step in this analysis is to depict the relationship between symptoms and disease. Consider:

$$P(D) = P(S_y) \cdot P(D/S_y) \tag{19}$$

where

$P(D)$ = the probability of occurance of disease,

$P(S_y)$ = probability of the occurance of a disease symptom, and

$P(D/S_y)$ = probability of the occurance of a disease given the presence of a symptom.

This equation illustrates that the probability of a disease occuring is the probability of having a symptom related to that disease multiplied by the probability of having the disease given that symptom.

As is evident from the analysis presented in the previous section, one of the determinants of disease symptoms is air pollution. Therefore, the probability of incuring a disease symptom, and the resultant probability of incuring the disease, is conditional upon a given level of air pollution. In this context equation (19) becomes:

$$P(D/\tilde{P}_o) = P(S_y/\tilde{P}_o).P\{D/Sy), \qquad (20)$$

where P_o is some given level of air pollution. Note that the probability of disease given a symptom is assumed independent of the pollution level.

For a change in the given level of air pollution, we observe:

$$P(D/\tilde{P}_1) - P(D/\tilde{P}_o) = [P(S_y/\tilde{P}_1) - P(Sy/Fe)] \cdot P(D/S_y); \qquad (21)$$

where $\tilde{P}_1$ is a new level of pollution. This implies that:

$$\Delta P(D/\Delta\tilde{P}) = \Delta P(S_y/\Delta\tilde{P}).P(D/S_y). \qquad (22)$$

Equation (22) illustrates that, as a result of a change in the level of air pollution, the change in the probability of incurring a disease is equal to the change in the probability of incuring a symptom multiplied by the associated probability of incuring a disease given the symptom.

From this analysis, the expected cost of disease can be defined as:

$$E(C_D/\tilde{P}_o) = P(D/\tilde{P}_o).CD, \qquad (23)$$

where C_D is the cost of disease. A change in the expected cost given a change in the pollution level is given by:

$$\Delta E(C_D/\Delta\tilde{P}) = \Delta P(D/\Delta\tilde{P}) \cdot C_D. \qquad (24)$$

Substitution of equation (22) into equation (24) yields:

$$\Delta E(C_D/\Delta\tilde{P}) = CD \cdot \Delta P(S_y/\Delta\tilde{P}).P(D/S_y). \qquad (25)$$

Equation (25) represents the change in the expected cost of disease given a unit change in the level of pollution for each symptom. The change in expected costs for each symptom can now be summed over the diseases to evaluate the total change in the expected cost of a symptom from a unit change in the pollution level.

Note that in the contex of the above analysis, an individual who has a disease symptom faces three possible states of the world. A symptom may exist and the individual has a disease or the symptom may exist without the corresponding presense of a disease. Further, since only certain rather specific diseases are considered in this analysis, it is possible that the individual who has a symptom does not have one of the diseases considered. The following diagram illustrates the possible situations:

Symptom exists ⇒ either/or
- No disease ⇒ zero economic cost
- Disease ⇒ and/or
  - Disease specific to study ⇒ positive economic cost (considered)
  - Other diseases ⇒ positive economic cost (not considered)

Only the upper half of the bottom chain is considered in the definition of economic costs in this analysis. Therefore, this study only concentrates on the economic costs of a few diseases. Economic costs of other diseases are not considered.

One possible source of distortion in this analysis arises due to the fact that the economic costs incurred by a person who has several different diseases simultaneously is probably lower than the simple summation of economic costs from the individual diseases. In this aspect, medical costs are lower for an individual suffering from. several diseases than for several individuals suffering from one disease. This arises due to the fact that the same treatment procedures may apply to many diseases and that some costs, such as office calls, hospitalization, and loss of work time are relatively fixed once a disease is incurred. These costs tend to remain nearly the same whether one or several diseases are treated in the same individual.

Nine different diseases were used as representative of the circulatory and respiratory diseases which have these symptoms. Although there are many other diseases which are related to these symptoms, the inability to acquire data on alternate diseases prevented their use in this study.

The expected economic costs associated with these nine diseases were taken from alternative sources and adjusted to per case estimates (14)(15)(16). The total economic cost of a disease per case is the sum of the direct, indirect and expected mortality costs. Per case adjustments were made using morbidity and mortality rates. Table 12 presents the per case annual economic costs of each disease by type of expenditure. For example, the estimated expected average cost to an individual from having ichemic heart disease is $7,388.11 per year in 1981 dollars. Of this amount, $3,422 are direct expenditures which consist of hospital expenditures, nursing home fees and expenditures on physician services and prescriptions. Indirect costs, loss of work time due to illness is $3,720. The rest of the total expected cost is made up of the expected loss of earnings due to death. Expected lost earnings of the individual are discounted present values calculated with an 8 percent discount rate.

TABLE 6.12 ESTIMATED ANNUAL PER CASE EXPECTED COST OF DISEASES, BY TYPE OF DISEASE, IN 1969 DOLLARS

	Direct[a] cost	Indirect[b] cost	Expected[c] Mortality cost	Total[d] Expected cost	Number of Deaths/ Year	Prevalence/ Year (Thousands)
Respiratory Diseases[e]						
Chronic Bronchitis	$57 (154)	$30 (81)	$.90 (2.45)	$87.90 (237.45)	5,305	6,526
Bronchiectasis	198 (537)	60 (163)	.25 (.68)	258.25 (700.68)	1,476	116
Emphysema	130 (352)	344 (932)	3.82 (10.35)	477.82 (1294.35)	20,873	1,313
Chronic Interstitial Pneumonia	62 (168)	-	9.96 (26.98)	71.96 (194.98)	4,218	403
Heart Diseases[f]						
Ischemic Disease	1391 (3422)	1512 (3720)	100.05 (246.11)	2931.05 (7,388.11)	669,829	1,333
Rheumatic Fever and Rheumatic Heart Disease	291 (716)	407 (1001)	3.44 (8.47)	701.44 (1725.47)	15,432	327
Cardiomyopathy	15 (37)	96 (236)	3.66 (8.99)	114.66 (281.99)	17,753	1,560
Arrhythmias	325 (800)	139 (342)	1.49 (3.66)	465.49 (1145.66)	7,298	389
Cardiac Failure	2736 (6731)	418 (1028)	1.67 (4.12)	3155.67 (7763.12)	11,388	113

TABLE 12 (continued)

- = Insufficient data

[a]For heart disease direct costs = hospital expenditures + nursing home expenditures + expenditures on physician services. For respiratory disease direct costs = hospital expenditures + nursing home expenditures + expenditures on physician services + expenditures on prescriptions.

[b]Indirect cost = loss of earnings due to illness or disability.

‘Expected mortality cost = expected loss of earnings due to death = (probability of death from disease) .(loss of earnings due to death). For respiratory disease a 6% discount rate is used, for heart disease an 8% discount rate is utilized.

[d]Expected total cost = direct + indirect cost + expected mortality cost.

‘Heart disease data is in 1969 dollars and utilized 1969 and 1970 data. The figures in ( ) are adjusted to 1981 dollars.

[f]Respiratory data is in 1967 dollars and utilized 1967 and 1970 data. The figures in ( ) are adjusted for 1981 dollars.

References:
1. Acton, Jan Paul, “Measuring the Social Impact of Heart and Circulatory Disease Programs: Preliminary Framework and Estimates,” Rand Corp. R-1697-NHLI, April 1975.

2. U.S. National Heart and Lung Institute, “Respiratory Diseases: Task Force Report on Problems, Research Approaches, Needs,” DHEW Pub. No. (NIH) 76-432, pp. 205-243, October 1972.

3. Department of Health, Education and Welfare, National Center for Health Statistics, “Prevalence of Selected Chronic Respiratory Conditions,” DHEW Pub. No. (HRA) 74-1511, Series 10, 84, 1970.

Expected values are a necessary component of the total cost of a disease since all individuals who have a disease do not necessarily die from the disease. This necessitates the use of an expected cost of mortality in the calculations. This number represents the loss of earnings due to death multiplied by the disease specific mortality rate. The mortality rate is the probability that an individual will die from the disease in question. Therefore, in this context the per case expected cost of disease becomes:

$$E(C_D) = d + i + E(m) \tag{26}$$

where

$E(C_D)$ = the expected cost of disease,

d = direct costs,

i = indirect costs, and

$E(m)$ = Probability of Death . Loss of Earnings due to Death = The Per Case Expected Cost of Death

Ideally, to depict the probability of death in this study, a mortality rate should be used which is conditional upon the presence of disease symptoms. However, since this information was unobtainable, per capita mortality rates derived for the society (of the U.S.) as a whole were used as a proxy. These rates are presented in the first column of Table 13.

Use of the societal mortality rate instead of a rate conditional on the existence of disease symptoms induced a downward bias to cost estimates. This is due to the fact that death rates due to disease are undoubtable higher in persons who already experience disease symptoms than in the society as a whole.

Note now that equation 25 must be modified to include the expected cost of disease. Equation 25 becomes:

$$\Delta E(C_D/\Delta\tilde{P}) = E(C_D).\Delta P(S_y/\tilde{P}).P(D/S_y) \tag{27}$$

Equation 27 forms the basis for derivation of cost savings due to reductions in the level of air pollution presented in this study. The first term on the left hand side, the per case expected cost of disease, is presented in Table 12. The second term, the change in the probability of incurring a disease symptom given a unit change in the level of air pollution, is simply the regression coefficient on air pollution variables which are presented in Section IV. The third and final term necessary to calculate the change in costs arising from a reduction in air pollution, the probability of disease given a symptom, is proxied in this analysis via the societal prevalence rate for the disease in question.

Again, as in the above discussion on mortality, use of the societal prevalence rate for a disease as a proxy for the incidence of that disease in individuals who already show evidence of symptoms will introduce a down-

TABLE 6.13 PER CAPITA PREVALENCE AND MORTALITY RATES OF SPECIFIC DISEASES IN THE UNITED STATES

	Mortality Rate	Prevalence Rate
Respiratory Diseases[a]		
Chronic Bronchitis	.00004	.03185
Bronchiectasis	.00001	.00057
Emphysema	.00018	.00641
Chronic Interstitial Pneumonia	.00004	.00197
Heart Diseases[b]		
Ischemic Disease	.00330	.00658
Rheumatic Fever and Rheumatic Heart Disease	.00007	.00161
Cardiomyopathy	.00009	.00769
Arrythmias	.00004	.00192
Cardiac Failure	.00006	.00056

[a]Based on number of deaths in 1967 and prevalence in 1970 from Table 12 and a U.S. population of 119,118,000 in 1967, U.S. Department of Commerce, Current Population Reports: Population Estimates and Projections, pg. 12, July 31, 1982, and a U.S. population 204,879,000 in 1970, Ibid., U.S. Department of Commerce, pg. 11, December 1972.

[b]Ibid., prevalence and deaths in 1969 from Table 12 and U.S. population of 202,677,000 in 1969, Ibid., U.S. Department of Commerce, pg. 11, December 1972.

ward bias to the results. This occurs due to the fact that, at the margin, the change in the probability of incurring a disease given a change in a symptom will be larger than the corresponding change in the incidence rate of that disease in the society as a whole. Societal prevalence rates for the nine diseases considered in this analysis are presented in the second column of Table 13.

Per capita estimates of the change in expected cost of disease given a unit change in the pollution level, derived via equation 27, are presented in the first column of Table 14. To derive these estimates, information from Table 12, Table 13 and the regression tables of Section IV are used. Note that these costs are presented by symptom and that they are adjusted to reflect 1981 dollars.

These results can be summed over diseases to yield per case estimates of the total cost of symptom given a unit change in air pollution. The last column of Table 14 presents these results. Note that not all symptoms apply to each disease and vice versa.

Table 15 presents estimates of cost of benefits in relation to unit changes in pollution levels. For extrapolative purposes, change in expected cost is assumed to be independent of the initial level of pollution. Intuitively, one would expect an increasing average relationship between the costs (benefits) incurred from a pollution increase (decrease) and the initial pollution level. This is illustrated garphically in Figure 5. If the initial level is P_1 and a change in the pollution level occurs bringing society to a level of P_1^o the benefits received are B_1. Now if the initial level is $\tilde{P}_1$ and a reduction in pollution of the same amount as above occurs, $\Delta\tilde{P}$, the benefits received will be less than B_1 and are equal to $\tilde{B}_1$. However, it has been demonstrated that rather than increasing average benefits for increasing initial levels of pollution, there may be decreasing average benefits (17). Due to uncertainty surrounding the actual relationship, a linear relationship between pollution changes and economic costs is assumed to hold for purposes of extrapolating the results to larger pollution changes.

In order to derive estimates of total United States cost savings due to a reduction in air pollution, a 30 percent improvement in mean air quality is assumed. These results are presented in Table 16. Total cost savings are presented, by symptom, for males between the ages of 55 and 64 and for the total population in the United States. Male members of the U.S. population between 55 and 64 years of age most closely represent the twins sample as characterized by 1980 census data. A more proper characterization of the twins data set is to include all males 55 to 65 years of age in 1981. However, due to limitations in census data, this categorization is not possible. Approximately 10,178,000 males were in this age group in 1980. At that time, the total U.S. population was about 226,505,000.

A 30 percent reduction in average maximum 24 hour concentration of SO_2 and TSP implies that mean levels of SO_2 will be reduced by 14.88 $\mu g/m^3$ and TSP will be reduced by 38.86 $\mu g/m^3$. Therefore, total cost savings, per symptom, can be calculated via the following formula:

TABLE 6.14 THE CHANGE IN THE TOTAL ANNUAL pER CAPITA EXPECTED COST OF A SYMPTOM DUE TO A UNIT CHANGE IN THE POLLUTION LEVEL, BY SYMPTOM AND DISEASE[a]

Symptom	Disease	$\Delta E(C_D/\Delta\tilde{P})$	Change in Total Cost of Symptom Given a Unit Change in the Pollution Level	
			TSP	SO_2
Cough	Chronic Bronchitis	.00045	.00391	
	Bronchiecstasis	.00002		
	Emphysema	.00050		
	Chronic Interstitial Pneumonia	.00001		
	Ischemic Heart Disease	.00292		
Shbr	Chronic Bronchitis	.00030	. 00308[b]	
	Bronchiecstasis	.00002		
	Emphysema	.00033		
	Chronic Interstitial Pneumonia	.00002		
	Ischemic Heart Disease	.00195		
	Rheumatic Heart Disease	.00011		
	Cardiomyopathy	.00009		
	Arrhythmias	.00009		
	Coronary Heart Attack	.00017		
Chpn	Chronic Bronchitis	.00076		.00693
	Bronchiecstasis	.00004		
	Emphysema	.00083		
	Ischemic Heart Disease	.00486		
	Cardiomyopathy	.00022		
	Arrhythmias	.00022		

TABLE 14 (continued)

Svchpn	Ischemic Heart Disease	.00049	.00053[b]
	Cardiac Failure	.00004	
Corn	Cardiac Failure	.00087	.00087

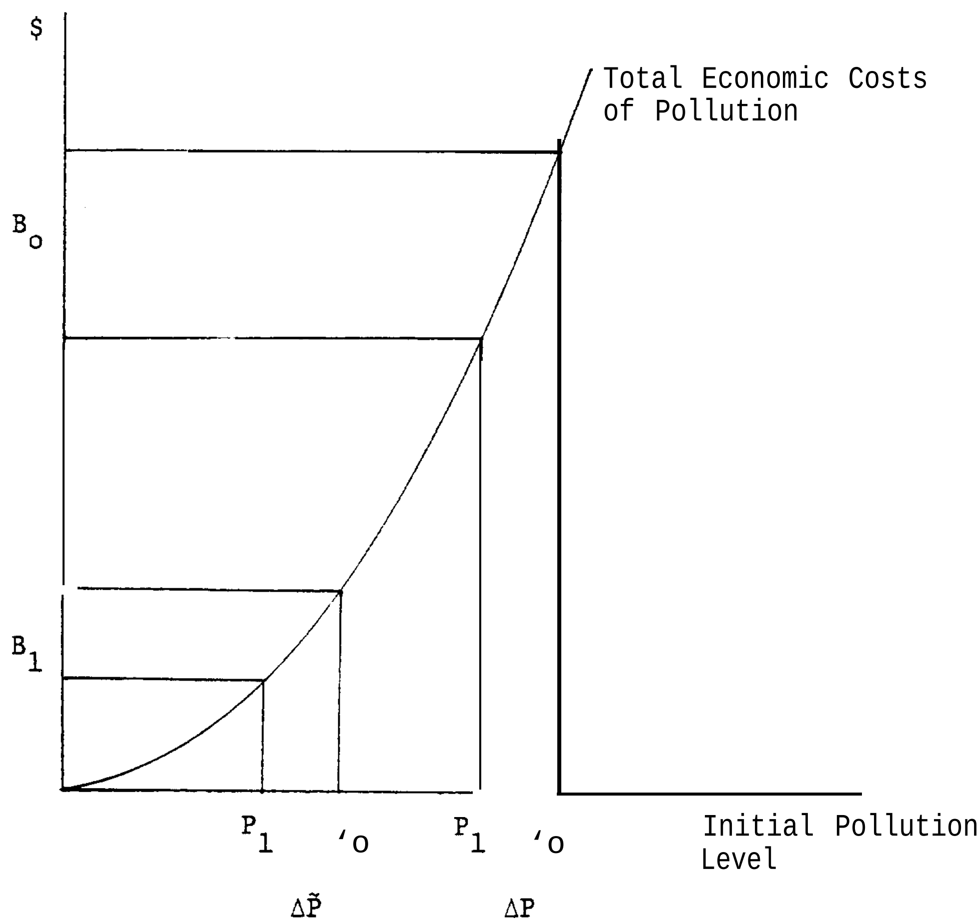
[a] values are reported only if the regression coefficient has a positive sign

[b] the coefficients used from the regression analysis to calculate these figures were not significant at the 90 percent level

TABLE 6.15 CHANGE IN PER CAPITA ANNUAL EXPECTED COST OF SYMPTOM GIVEN A CHANGE IN THE POLLUTION LEVEL

Symptom	Unit Change in the Pollution Level[a] $1 \mu g/m^3$
Cough	.00391
Shortness of Breath	.00308
Chest Pain	.00693
Severe Chest Pain	.00053
Cardiac Failure	.00087

[a]TSP is used for all the symptoms except for chest pain where SO_2 is used.



Where $\Delta\tilde{P} = \Delta P$

Figure 6.5 Measuring benefits from pollution reduction assuming increasing costs of pollution

TABLE 6.16 TOTAL COST SAVINGS, BY SYMPTOM, FOR A 30 PERCENT IMPROVEMENT IN U.S. AIR QUALITY IN 1981 DOLLARS[a]

Symptom	Total for males between 55-64 years of age[b]	Total U.S. Population[b]
Cough	$1,546,000	$34,416,000
Shortness of Breath	1,218,000	27,110,000
Chest Pain[c]	1,050,000	23,357,000
Severe Chest Pain	210,000	4,665,000
Cardiac Failure	344,000	7,658,000
TOTAL	4,368,000	97,206,000

[a] mean values for SO_2 and TSP were used as initial values

[b] 1980 census of population data

[c] SO_2 is the air pollution variable used here and TSP is used for all other symptoms

$$\text{Total Cost Saving} = \text{Population} \cdot \text{Reduction in Air Pollution} \cdot \text{Per Case Cost of Symptom}$$

A 30 percent reduction in TSP is assumed for all symptoms except for chest pain where a 30 percent reduction in SO_2 is assumed.

Summation over the five symptoms yields an overall measure of the health benefits of air quality improvement. Note that for the age group nearest to the twins sample, total cost savings from disease is over $4 million. If these results are extrapolated to the entire U.S. population, a savings of nearly $100 million is incurred.

In order to compare this result to Ostro (1982) (18) and Crocker et.al. (1979) (19), it is necessary to exclude the cost savings arising from a reduction in SO_2 and only consider the costs savings arising from a reduction in total suspended particulate. Cost savings are reduced by $23,357,000 to $73,849,000 (in 1981 dollars) when only a 30 percent reduction in TSP is considered.

Ostro (1982) estimated that a 19 percent reduction in TSP will yield an urban benefit by reducing the number of work loss days by a range of 3 to 78 million. If a daily average wage of $46.00 is assumed for 1981, the range of damages in Ostro's analysis becomes $138 million to 3.588 billion.

Crocker et.al. (1979) analyzed the urban benefits of reduced mortality. Using the mean concentration of TSP in a sixty-city sample, they estimated the average reduction in risk of pneumonia mortality for a 60 percent reduction in particulate. Urban benefits of reduced mortality due to a 60 percent reduction in the level of total suspended particulate were estimated to be within a range of 5.4 to 16.7 billion dollars (adjusted to 1981 dollars).

In comparing the results presented in this paper to these other studies, one notes that the symptom sensitive analysis utilized here yields a lower bound. Only the lower end of Ostro’s range is comparable with the results of this paper. Crocker et.al. estimates are much larger than the benefits estimated in either this study or Ostro's.

However, one can note that Ostro's results, which were calculated across all diseases, represent a marginal representation of work loss days. The indirect costs of disease presented in the d'Arge et.al. analysis were based on average work loss days due to a few specific diseases. In this aspect we would fully expect marginal work loss days to be larger than average work loss days because days lost increase as pollution increases.

Further, in considering the Crocker et.al. results, it must be realized that their results were based on the population as a whole while the d'Arge et.al. results were calibrated to a very specific sample of the population. At the time health statistics were collected for the twins data set, the group ranged in age from 41 to 51 years. In this context, the twins sample represented a fairly healthy segment of society. The Crocker et.al. sample included many older individuals whom we would expect would be more effected by air pollution. Therefore, the Crocker et.al. result should exceed the d'Arge et.al. results in magnitude.

Finally, one should not forget the impact of the use of societal prevalence and death rates to proxy rates in individuals who exhibit disease symptoms in the d'Arge et.al. analysis. This phenomena will also result in the d'Arge et.al. results being lower bounds.